

ON THE SCALING WINDOW OF MODEL RB

CHUNYAN ZHAO¹, KE XU², AND ZHIMING ZHENG³

ABSTRACT. This paper analyzes the scaling window of a random CSP model (i.e. model RB) for which we can identify the threshold points exactly, denoted by r_{cr} or p_{cr} . For this model, we establish the scaling window $W(n, \delta) = (r_-(n, \delta), r_+(n, \delta))$ such that the probability of a random instance being satisfiable is greater than $1 - \delta$ for $r < r_-(n, \delta)$ and is less than δ for $r > r_+(n, \delta)$. Specifically, we obtain the following result

$$W(n, \delta) = (r_{cr} - \Theta(\frac{1}{n^{1-\varepsilon} \ln n}), r_{cr} + \Theta(\frac{1}{n \ln n})),$$

where $0 \leq \varepsilon < 1$ is a constant. A similar result with respect to the other parameter p is also obtained. Since the instances generated by model RB have been shown to be hard at the threshold, this is the first attempt, as far as we know, to analyze the scaling window of such a model with hard instances.

1. INTRODUCTION

The Constraint Satisfaction Problem(CSP), originated from artificial intelligence, has become an important and active field of statistical physics, information theory and computer science. The CSP area is very interdisciplinary, since it embeds ideas from many research fields, like artificial intelligence, databases, programming languages and operation research. A constraint satisfaction problem consists of a finite set $U = \{u_1, u_2, \dots, u_n\}$ of n variables, each u_i associated with a domain of values D_i , and a set of constraints. Each of the constraints $C_{i_1 i_2 \dots i_k}$ is a relation, defined on some subset $\{u_{i_1}, u_{i_2}, \dots, u_{i_k}\}$ of n variables, called its scope, denoting their legal tuples of values. A *solution* to a CSP is an assignment of a value to each variable from its domain such that all the constraints of this CSP are satisfied. A constraint is said to be *satisfied* if the tuple of values assigned to the variables in this constraint is a legal one. A CSP is called *satisfiable* if and only if it has at least one solution. The task of a CSP is to find a solution or to prove that no solution exists.

Given a CSP, we are interested in polynomial-time algorithms, that is, algorithms whose running time is bounded by a polynomial in the number of variables. Cook's Theorem[2] asserts that satisfiability is NP-complete and at least as hard as any problem whose solutions can be verified in polynomial time. Most of the interesting CSPs are NP-complete problems. We know that k -SAT problem is a canonical version of the CSPs, in which variables can be assigned the value True or False(called Boolean variables). A lot of efforts have been devoted to k -SAT and it is widely believed that no efficient algorithm exists for k -SAT. However,

Key words and phrases. Constraint Satisfaction Problem; Model RB; Satisfiability; Phase Transition; Scaling Window.

Supported in part by National 973 Program of China(Grant No. 200532CB1902) and NSFC(Grant Nos. 60473109 and 60403003).

it is shown that most instances of k -SAT can be solved efficiently, so perhaps genuine hardness is only present in a tiny fraction of all instances. In 1990s, a remarkable progress[3, 13, 11, 12] was made that the really difficult instances is related to phase transition phenomenon, as suggested in the pioneering work of Fu and Anderson[6]. The study of phase transitions has attracted much interest subsequently[9, 12].

In recent years, random k -SAT has been well studied both from theoretical and algorithmic point of views. If $k = 2$ then it is known that there is a satisfiability threshold at $\alpha_c = 1$ (here α represents the ratio of clauses m to variables n), below which the probability of a random instance being satisfiable tends to 1 and above which it tends to 0 as n approaches infinity[4]. This was sharpened in[8, 14]. Random 2-SAT is now pretty much understood. However, for $k \geq 3$, the existence of the phase transition phenomenon has not been established, not even the exact value of the threshold point[1, 10].

To gain a better understanding of how the phase transition scales with problem size, the finite-size scaling method has been introduced from statistical mechanics[11, 7]. We use finite-size scaling, a method from statistical physics in which observing how the width of a transition narrows with increasing sample size gives direct evidence for critical behavior at a phase transition. Finite-size scaling is the study of changes in the transition behavior due to finite-size effects, in particular, broadening of the transition region for finite n . More precisely, for $0 < \delta < 1$, let $r_-(n, \delta)$ be the supremum over r such that the probability of a random CSP instance being satisfiable is at least $1 - \delta$, and similarly, let $r_+(n, \delta)$ be the infimum over r such that the probability of a random CSP instance being satisfiable is at most δ . Then, for r within the scaling window $W(n, \delta) = (r_-(n, \delta), r_+(n, \delta))$ the probability is between δ and $1 - \delta$. And for all δ , $|r_+(n, \delta) - r_-(n, \delta)| \rightarrow 0$ as $n \rightarrow \infty$. For random 2-SAT, it has been determined that the scaling window is $W(n, \delta) = (1 - \Theta(n^{-1/3}), 1 + \Theta(n^{-1/3}))[2]$.

Model RB is a random CSP model proposed by Xu and Li to overcome the trivial insolubility of standard CSP models[16]. For this model, we can not only establish the existence of phase transitions, but also pinpoint the threshold points exactly, denoted by r_{cr} or p_{cr} . Moreover, it has been proved that almost all instances of model RB have no tree-like resolution proofs of less than exponential size [16]. This implies that unlike random 2-SAT, model RB can be used to generate hard instances, which has also been confirmed by experiments[7]. Motivated by the work on the scaling window of random 2-SAT, in this paper, we study the scaling window of model RB and obtain that $W(n, \delta) = (r_{cr} - \Theta(\frac{1}{n^{1-\epsilon} \ln n}), r_{cr} + \Theta(\frac{1}{n \ln n}))$. And we also obtain similar results about the other control parameter p .

The main contribution of this paper is not to present new methods for computing the scaling window, but to show that for an interesting model with hard instances (i.e. model RB), not only can the threshold points be located exactly, but also the scaling window can be determined using standard methods. This means that hopefully, more mathematical properties about the threshold behavior of model RB can be obtained in a relatively easy way, which will help to shed light on the phase transition phenomenon in NP-complete problems. The rest of the paper is organized as follows. In the next section, we will give a brief introduction about model RB. The main results of this paper and their proofs will be given in Section 3 and Section 4 respectively. Finally, we will conclude in Section 5.

2. MODEL RB

We can pinpoint the threshold location for model RB proposed by Xu and Li[16]. The way of generating random instances for model RB is:

- (1). Given a set U of n variables, select with repetition $m = rn \ln n$ random constraints. Each random constraint is formed by selecting without repetition k of n variables, where $k \geq 2$ is an integer.
- (2). Next, for each constraint we select uniformly at random without repetition $q = p \cdot d^k$ illegal tuples of values, i.e., each constraint contains exactly $(1 - p) \cdot d^k$ legal ones, where $d = n^\alpha$ is the domain size of each variable and $\alpha > 0$ is a constant.

In this paper, the probability of a random CSP instance being satisfiable is denoted by $\Pr(\text{Sat})$. It is proved that for model RB the phase transition phenomenon occurs at $r_{cr} = -\frac{\alpha}{\ln(1-p)}$ or $p_{cr} = 1 - e^{-\frac{\alpha}{r}}$ as n approaches infinity[16]. More precisely, we have the following two theorems.

Theorem 2.1[16] *Let $r_{cr} = -\frac{\alpha}{\ln(1-p)}$. If $\alpha > \frac{1}{k}$, $0 < p < 1$ are two constants and k, p satisfy the inequality $k \geq \frac{1}{1-p}$, then*

$$\lim_{n \rightarrow \infty} \Pr(\text{Sat}) = 1 \quad \text{when } r < r_{cr},$$

$$\lim_{n \rightarrow \infty} \Pr(\text{Sat}) = 0 \quad \text{when } r > r_{cr}.$$

Theorem 2.2[16] *Let $p_{cr} = 1 - e^{-\frac{\alpha}{r}}$. If $\alpha > \frac{1}{k}$, $r > 0$ are two constants and k, α satisfy the inequality $ke^{-\frac{\alpha}{r}} \geq 1$, then*

$$\lim_{n \rightarrow \infty} \Pr(\text{Sat}) = 1 \quad \text{when } p < p_{cr},$$

$$\lim_{n \rightarrow \infty} \Pr(\text{Sat}) = 0 \quad \text{when } p > p_{cr}.$$

3. MAIN RESULTS

Our main results are the following two theorems.

Theorem 3.1 For all sufficiently small $\delta > 0$, there exist $r_-(n, \delta)$ and $r_+(n, \delta)$ such that the following holds:

$$\Pr(\text{Sat}) > 1 - \delta, \quad \text{when } r < r_-(n, \delta);$$

$$\Pr(\text{Sat}) < \delta, \quad \text{when } r > r_+(n, \delta),$$

where $r_-(n, \delta) = r_{cr} - \Theta(\frac{1}{n^{1-\varepsilon} \ln n})$, $r_+(n, \delta) = r_{cr} + \Theta(\frac{1}{n \ln n})$. So that the scaling window of model RB is

$$W(n, \delta) = (r_{cr} - \Theta(\frac{1}{n^{1-\varepsilon} \ln n}), r_{cr} + \Theta(\frac{1}{n \ln n})).$$

It is easy to see that $|r_+(n, \delta) - r_-(n, \delta)| \rightarrow 0$, as $n \rightarrow \infty$.

Theorem 3.2 For all sufficiently small $\delta > 0$, there exist $p_-(n, \delta)$ and $p_+(n, \delta)$

such that the following holds:

$$\begin{aligned} Pr(Sat) &> 1 - \delta, \quad \text{when } p < p_-(n, \delta); \\ Pr(Sat) &< \delta, \quad \text{when } p > p_+(n, \delta), \end{aligned}$$

where $p_-(n, \delta) = p_{cr} - \Theta(\frac{1}{n^{1-\varepsilon} \ln n})$, $p_+(n, \delta) = p_{cr} + \Theta(\frac{1}{n \ln n})$. So that the scaling window of Model RB is

$$W(n, \delta) = (p_{cr} - \Theta(\frac{1}{n^{1-\varepsilon} \ln n}), p_{cr} + \Theta(\frac{1}{n \ln n})).$$

It is not difficult to see that $|p_+(n, \delta) - p_-(n, \delta)| \rightarrow 0$, as $n \rightarrow \infty$.

Remark 3.1 If $n \rightarrow \infty$, then $r_+(n, \delta), r_-(n, \delta) \rightarrow r_{cr}$, $p_+(n, \delta), p_-(n, \delta) \rightarrow p_{cr}$. For every sufficiently small δ , Theorem 3.1 and Theorem 3.2 hold. So we can obtain

$$\begin{aligned} \lim_{n \rightarrow \infty} Pr(Sat) &= 1 \quad \text{when } r < r_{cr} \quad \text{or } p < p_{cr}, \\ \lim_{n \rightarrow \infty} Pr(Sat) &= 0 \quad \text{when } r > r_{cr} \quad \text{or } p > p_{cr}. \end{aligned}$$

This is the result of Xu and Li[16].

4. PROOF OF THE RESULTS

To prove the main results, we need the following lemmas.

Lemma 4.1 Let $c = \alpha + 1 - r_{cr}kp$, then $c < 1$.

Proof We know that $r_{cr} = -\frac{\alpha}{\ln(1-p)}$, then

$$\begin{aligned} c &= \alpha + 1 + \frac{\alpha kp}{\ln(1-p)} \\ &= 1 + \frac{\alpha[kp + \ln(1-p)]}{\ln(1-p)} \end{aligned}$$

Assume that $f(p) = kp + \ln(1-p)$, hence we have $f'(p) = -\frac{1}{1-p} + k$.

By the condition of Theorem 2.1, we have $k \geq \frac{1}{1-p}$, hence $f'(p) \geq 0$. That is $f(p)$ is a monotone increasing function.

So $f(p) > f(0)$, that is $kp + \ln(1-p) > 0$. It is obvious that $\ln(1-p) < 0$ because of $0 < p < 1$. And $\alpha > \frac{1}{k}$ is a constant.

Hence $\frac{\alpha[kp + \ln(1-p)]}{\ln(1-p)} < 0$.

Therefore, it is proved that $c = 1 + \frac{\alpha[kp + \ln(1-p)]}{\ln(1-p)} < 1$.

Lemma 4.2 Let $c = \alpha + 1 - rkp_{cr}$, then $c < 1$.

Proof We know that $p_{cr} = 1 - e^{-\frac{\alpha}{r}}$, so

$$\begin{aligned} c &= \alpha + 1 - rk(1 - e^{-\frac{\alpha}{r}}) \\ &= 1 - r[-\frac{\alpha}{r} + k(1 - e^{-\frac{\alpha}{r}})] \end{aligned}$$

Let $-\frac{\alpha}{r} = x$, then $x \in (-\infty, 0)$. Suppose $h(x) = x + k(1 - e^x)$, then $h'(x) = 1 - ke^x$.

By the condition of Theorem 2.2, $ke^x = ke^{-\frac{\alpha}{r}} \geq 1$, hence $h'(x) \leq 0$. That is $h(x)$ is a monotone decreasing function.

So $h(x) > h(0)$, that is $h(x) > 0$. And $r > 0$ is a constant, hence it is proved that $c = 1 - r[-\frac{\alpha}{r} + k(1 - e^{-\frac{\alpha}{r}})] < 1$.

Proof of Theorem 3.1 Let N denote the number of satisfying assignments for a random CSP instance, we can obtain that

$$\begin{aligned} E(N) &= d^n (1-p)^{rn \ln n} \\ (4.1) \quad &= n^{\alpha n} (1-p)^{rn \ln n} \end{aligned}$$

Assume that $E(N) < \delta$, by (1) we get

$$(4.2) \quad [\alpha + r \ln(1-p)]n \ln n < \ln \delta$$

$$(4.3) \quad \alpha + r \ln(1-p) < \frac{\ln \delta}{n \ln n}$$

$$(4.4) \quad r > -\frac{\alpha}{\ln(1-p)} + \frac{\ln \delta}{n \ln n \ln(1-p)} = r_{cr} + \frac{\ln \delta}{n \ln n \ln(1-p)}$$

Using the Markov inequality $\Pr(\text{Sat}) \leq E(N)$, we get $\Pr(\text{Sat}) < \delta$ for

$$(4.5) \quad r > r_{cr} + \Theta\left(\frac{1}{n \ln n}\right).$$

Here note that $f = \Theta(g)$ represents there exist two finite constants $c_1 > 0$ and $c_2 > 0$ such that $c_1 < f/g < c_2$.

In the following, we use Cauchy inequality $\Pr(\text{Sat}) \geq \frac{E^2(N)}{E(N^2)}$ to prove when $r < r_{cr} + \Theta(\frac{1}{n \ln n})$, we have $\Pr(\text{Sat}) > 1 - \delta$.

In the remaining part of the paper, the expression of $E(N^2)$ will play an important role in the proof of the main results. The derivation of this expression can be found in [16]. For the convenience of the reader, we give an outline of it as follows.

Definition 4.1 Let $\langle t_i, t_j \rangle$ represents an ordered assignment pair to the n variables in U , which satisfies a CSP instance if and only if both t_i and t_j satisfy the CSP instance. And $P(\langle t_i, t_j \rangle)$ denotes the probability of $\langle t_i, t_j \rangle$ satisfying a CSP instance.

Definition 4.2 The similarity number S of an assignment pair $\langle t_i, t_j \rangle$ is the number of variables t_i and t_j take the identical values. It is obvious that $0 \leq S \leq n$, and let $s = \frac{S}{n}$. Let A_S be the set of assignments whose similarity number is equal to S .

We can get the expression of $E(N^2)$ is

$$\begin{aligned} |A_S|P(\langle t_i, t_j \rangle) &= \sum_{S=0}^n |A_S|P(\langle t_i, t_j \rangle) \\ &= d^n \binom{n}{S} (d-1)^{n-S} \left[\frac{\binom{d^k-1}{q}}{\binom{d^k}{q}} \cdot \frac{\binom{S}{k}}{\binom{n}{k}} + \frac{\binom{d^k-2}{q}}{\binom{d^k}{q}} \cdot \left(1 - \frac{\binom{S}{k}}{\binom{n}{k}}\right) \right]^{rn \ln n} \end{aligned}$$

First we need to estimate $E(N^2)$. We can rewrite the above equation as the following one

$$(4.6) \quad |A_S|P(\langle t_i, t_j \rangle) = \frac{E^2(N) \left[1 + \frac{p}{1-p} \left(s^k + \frac{g(s)}{n}\right)\right]^{rn \ln n}}{\left(1 - \frac{1}{n^\alpha}\right)^{n-ns} \left(\frac{1}{n^\alpha}\right)^{ns} \binom{n}{ns} \left(1 + O\left(\frac{1}{n}\right)\right)}$$

where $g(s) = \frac{k(k-1)(s^k - s^{k-1})}{2}$.

When n is sufficiently large, except $E^2(N)$, the dominant contribution to (4.6) comes from

$$\begin{aligned} f(s) &= \left(1 + \frac{p}{1-p} s^k\right)^{rn \ln n} \left(\frac{1}{n^\alpha}\right)^{ns} \\ (4.7) \quad &= e^{[r \ln(1 + \frac{p}{1-p} s^k) - \alpha s] n \ln n} \end{aligned}$$

We put $h(s) = r \ln(1 + \frac{p}{1-p} s^k) - \alpha s$ and focus on the function $h(s)$, differentiating $h(s)$ twice with respect to s we get

$$(4.8) \quad h''(s) = \frac{rkps^{k-2}[(k-1)(1-p) - ps^k]}{(1-p+ps^k)^2}$$

Applying the condition $k \geq \frac{1}{1-p}$, we get $(k-1)(1-p) - ps^k \geq 0$ on the interval $[0, 1]$, then $h''(s) \geq 0$. So $h(s)$ is a convex function. It is easy to see that $h(0) = 0$ and $h(1) = -r \ln(1-p) - \alpha$. So when $r < r_{cr} - \Theta(1/(n^{1-\varepsilon} \ln n))$, we have $h(1) \leq 0$. On the interval $0 < s < 1$, we get $h(s) < 0$. So there exist $0 < \delta_1 < 1$ and $0 < \delta_2 < 1$ such that when $r < r_{cr} - \Theta(1/(n^{1-\varepsilon} \ln n))$, $h(s)$ is mainly decided by the values $s \in [0, \delta_1] \cup [1 - \delta_2, 1]$. So we only need to consider those terms $s \in [0, \delta_1] \cup [1 - \delta_2, 1]$ to estimate (4.6). This is different from the proof in Xu and Li[16] for establishing the existence of phase transitions, where only those terms $s \in [0, \delta_1]$ were considered.

(i) $s \in [0, \delta_1]$

We can learn from Xu and Li[16] that

$$(4.9) \quad \sum_{s \in [0, \delta_1]} |A_S| P(\langle t_i, t_j \rangle) \leq E^2(N) (1 + O(\frac{1}{n}))$$

(ii) $s \in [1 - \delta_2, 1]$

It is easily known that if $s \in [1 - \delta_2, 1]$, we can obtain $s^k - s^{k-1} < 0$, thus $g(s) = \frac{k(k-1)(s^k - s^{k-1})}{2} < 0$. So we can get the following inequality

$$\begin{aligned} |A_S| P(\langle t_i, t_j \rangle) &\leq E^2(N) \left(1 + \frac{p}{1-p} s^k\right)^{rn \ln n} \\ &\quad \cdot \left(1 - \frac{1}{n^\alpha}\right)^{(n-ns)} \left(\frac{1}{n^\alpha}\right)^{ns} \binom{n}{ns} \left(1 + O\left(\frac{1}{n}\right)\right) \\ (4.10) \quad &= E(N) (1 - p + ps^k)^{rn \ln n} (n^\alpha - 1)^{n-ns} \binom{n}{ns} \left(1 + O\left(\frac{1}{n}\right)\right) \end{aligned}$$

When $s = 1 (S = n)$, we obtain

$$(4.11) \quad |A_S| P(\langle t_i, t_j \rangle) = E(N) (1 + O(\frac{1}{n}));$$

When $s = \frac{n-t}{n}$ ($S = n - t$), where $1 \leq t \ll n$. We can get that

$$\begin{aligned}
|A_S|P(\langle t_i, t_j \rangle) &\leq E(N) \cdot [1 - p + p(\frac{n-t}{n})^k](n^\alpha - 1)^t \binom{n}{t} \cdot (1 + O(\frac{1}{n})) \\
&\leq E(N) e^{-p[1 - (\frac{n-t}{n})^k] r n \ln n} (n^\alpha - 1)^t \binom{n}{t} \cdot (1 + O(\frac{1}{n})) \\
&\leq E(N) \frac{n^{(\alpha+1)t}}{n^{r n p t (\frac{k}{n} - O(\frac{1}{n^2}))}} (1 + O(\frac{1}{n})) \\
&= E(N) \frac{n^{(\alpha+1)t}}{n^{r k p t - O(\frac{1}{n})}} (1 + O(\frac{1}{n})) \\
(4.12) \quad &\leq E(N) (\frac{n^{\alpha+1+O(\frac{1}{n})}}{n^{r k p}})^t (1 + O(\frac{1}{n}))
\end{aligned}$$

When n is sufficiently large, let $c = \alpha + 1 - r_{cr} k p = \alpha + 1 + \frac{\alpha k p}{\ln(1-p)}$. Thus it is divided into two cases to discuss the value of c .

Case 1: $c < 0$.

When $s = \frac{n-1}{n}$, by (4.12) we can obtain

$$(4.13) \quad |A_S|P(\langle t_i, t_j \rangle) \leq E(N) \cdot n^c \cdot (1 + O(\frac{1}{n}))$$

When $s = \frac{n-2}{n}$, by (4.12) we have

$$(4.14) \quad |A_S|P(\langle t_i, t_j \rangle) \leq E(N) \cdot n^{2c} \cdot (1 + O(\frac{1}{n}))$$

... ..

So we can get

$$\begin{aligned}
\sum_{s \in [1-\delta_2, 1]} |A_S|P(\langle t_i, t_j \rangle) &\leq E(N)(1 + n^2 + n^{2c} + \dots) \cdot (1 + O(\frac{1}{n})) \\
(4.15) \quad &= E(N)(1 + O(n^c))
\end{aligned}$$

It is shown from (i) and (ii) that

$$\begin{aligned}
E(N^2) &= \sum_{S=0}^n |A_S|P(\langle t_i, t_j \rangle) \\
&= \sum_{s \in [0, \delta_1]} |A_S|P(\langle t_i, t_j \rangle) + \sum_{s \in [1-\delta_2, 1]} |A_S|P(\langle t_i, t_j \rangle) \\
(4.16) \quad &\leq E^2(N)(1 + O(\frac{1}{n})) + E(N)(1 + O(n^c))
\end{aligned}$$

Consequently, by the Cauchy inequality, we have

$$\begin{aligned}
Pr(Sat) &\geq \frac{E^2(N)}{E(N^2)} \geq \frac{E^2(N)}{E^2(N)(1 + O(\frac{1}{n})) + E(N)(1 + O(n^c))} \\
(4.17) \quad &> 1 - \delta
\end{aligned}$$

$$(4.18) \quad E(N) > \frac{1 - \delta + O(n^c)}{\delta - O(\frac{1}{n})}$$

Putting $\frac{1-\delta+O(n^c)}{\delta-O(\frac{1}{n})} = \vartheta$, hence we have

$$(4.19) \quad \alpha n \ln n + r n \ln n \ln(1-p) > \ln \vartheta$$

$$(4.20) \quad r < \frac{\ln \vartheta - \alpha n \ln n}{n \ln n \ln(1-p)} = -\frac{\alpha}{\ln(1-p)} + \frac{\ln \vartheta}{n \ln n \ln(1-p)}$$

So we obtain that

$$(4.21) \quad r < r_{cr} + \frac{\ln \vartheta}{n \ln n \ln(1-p)}$$

Thus when $r < r_{cr} + \Theta(\frac{1}{n \ln n})$, we have the result $Pr(Sat) > 1 - \delta$.

Case 2: $c \geq 0$.

When $1 \leq t \ll n$, by the right side of (4.10), we can get

$$(4.22) \quad \begin{aligned} & [1 - p + p(1 - \frac{t}{n})]^{rn \ln n} (n^\alpha - 1)^t \binom{n}{t} \\ &= n^{-rkpt + O(\frac{1}{n})} n^{\alpha t} (1 - \frac{1}{n^\alpha})^t \sqrt{2\pi n} \left(\frac{n}{t}\right)^t (1 + O(\frac{1}{n})) \\ &\leq \frac{\sqrt{2\pi n} \cdot n^{(\alpha+1+O(\frac{1}{n})-rkp)t}}{t^t} (1 + O(\frac{1}{n})) \end{aligned}$$

Now when n is sufficiently large, let $u_t = \frac{n^{(\alpha+1-rkp)t}}{t^t} = \frac{n^{ct}}{t^t}$. Then $u_t = e^{ct \ln n - t \ln t}$. If we put $\omega_t = ct \ln n - t \ln t$, we can get $\omega'_t = c \ln n - \ln t - 1$, then $\omega'_t = 0$ when $t = \frac{n^c}{e}$. And it is known that $0 \leq c < 1$ by Lemma 4.1. So $|A_S|P(\langle t_i, t_j \rangle)$ has the maximal value $\sqrt{2\pi n} \cdot e^{\frac{n^c}{e}}$ at the point of $t = \frac{n^c}{e}$. So we can have

$$(4.23) \quad \sum_{s \in [1-\delta_2, 1]} |A_S|P(\langle t_i, t_j \rangle) \leq E(N) \sqrt{2\pi n} \cdot e^{\frac{n^c}{e}} n (1 + O(\frac{1}{n}))$$

We use the Cauchy inequality

$$(4.24) \quad \begin{aligned} Pr(Sat) &\geq \frac{E^2(N)}{E(N^2)} \geq \frac{E^2(N)}{(E^2(N) + E(N) \sqrt{2\pi n} \cdot e^{\frac{n^c}{e}} n) (1 + O(\frac{1}{n}))} \\ &> 1 - \delta \end{aligned}$$

$$(4.25) \quad E(N) > \frac{1 - \delta + O(\frac{1}{n})}{\delta - O(\frac{1}{n})} \sqrt{2\pi n} \cdot e^{\frac{n^c}{e}} n$$

Let $\frac{\sqrt{2\pi}(1-\delta+O(\frac{1}{n}))}{\delta-O(\frac{1}{n})} = \lambda$, then we get

$$(4.26) \quad \alpha n \ln n + rn \ln n \ln(1-p) > \ln \lambda + \frac{n^c}{e} + \frac{3}{2} \ln n$$

$$(4.27) \quad \begin{aligned} r &< \frac{-\alpha n \ln n + \frac{n^c}{e} + \frac{3}{2} \ln n + \ln \lambda}{n \ln n \ln(1-p)} \\ &= -r_{cr} + \frac{1}{en^{1-c} \ln n \ln(1-p)} + \frac{3}{2n \ln(1-p)} + \frac{\ln \lambda}{n \ln n \ln(1-p)} \end{aligned}$$

So when $r < r_{cr} + O(\frac{1}{n^{1-c} \ln n})$, we have $Pr(Sat) > 1 - \delta$.

Combining the above cases, it is proved that the scaling window of model RB is

$$W(n, \delta) = (r_{cr} - \Theta(\frac{1}{n^{1-\varepsilon} \ln n}), \quad r_{cr} + \Theta(\frac{1}{n \ln n})),$$

where $\varepsilon = \frac{c+|c|}{2}$, $c < 1$ and it is obvious that $|r_{cr} + \Theta(\frac{1}{n \ln n}) - (r_{cr} - \Theta(\frac{1}{n^{1-\varepsilon} \ln n}))| \rightarrow 0$ ($n \rightarrow \infty$). Thus, we finish the proof of Theorem 3.1.

Remark 4.1 By Lemma 4.1, we claim that c increases with p and decreases with α . Therefore, when $0 \leq c < 1$, the convergence rate of $r_-(n, \delta)$ approaching r_{cr} decreases with p and increases with α .

Proof of Theorem 3.2 Similarly, we can also use (4.3) to obtain that

$$\begin{aligned}
 (4.28) \quad \ln(1-p) &< -\frac{\alpha}{r} + \frac{\ln \delta}{rn \ln n} \\
 p &> 1 - e^{-\frac{\alpha}{r} + \frac{\ln \delta}{rn \ln n}} \\
 &= 1 - e^{-\frac{\alpha}{r}} + e^{-\frac{\alpha}{r}} (1 - e^{\frac{\ln \delta}{rn \ln n}}) \\
 &= p_{cr} + e^{-\frac{\alpha}{r}} [1 - (1 + O(\frac{\ln \delta}{rn \ln n}))] \\
 (4.29) \quad &= p_{cr} + \Theta(\frac{1}{n \ln n})
 \end{aligned}$$

So when $p > p_{cr} + \Theta(\frac{1}{n \ln n})$, we have $Pr(Sat) < \delta$.

Similar to the proof of Theorem 3.1, when n is sufficiently large, let $c = \alpha + 1 - rkp_{cr}$. So by Lemma 4.2 we can also divide c into two cases, that is to say $c < 0$ and $0 \leq c < 1$. Therefore, we have the followings.

By (4.19), we can get

$$\begin{aligned}
 (4.30) \quad \ln(1-p) &> \frac{\ln \vartheta - \alpha n \ln n}{rn \ln n} = -\frac{\alpha}{r} + \frac{\ln \vartheta}{rn \ln n} \\
 p &< 1 - e^{-\frac{\alpha}{r} + \frac{\ln \vartheta}{rn \ln n}} \\
 &= 1 - e^{-\frac{\alpha}{r}} + e^{-\frac{\alpha}{r}} (1 - e^{\frac{\ln \vartheta}{rn \ln n}}) \\
 &= p_{cr} + e^{-\frac{\alpha}{r}} [1 - (1 + O(\frac{\ln \vartheta}{rn \ln n}))] \\
 (4.31) \quad &= p_{cr} - \Theta(\frac{1}{rn \ln n})
 \end{aligned}$$

By (4.26), we have

$$\begin{aligned}
 (4.32) \quad \alpha n \ln n + rn \ln n \ln(1-p) &> \ln \lambda + \frac{n^c}{e} + \frac{3}{2} \ln n \\
 p &< 1 - e^{-\frac{\alpha}{r} + \frac{\frac{n^c}{e} + \frac{3}{2} \ln n + \ln \lambda}{rn \ln n}} \\
 &= 1 - e^{-\frac{\alpha}{r}} + e^{-\frac{\alpha}{r}} (1 - e^{\frac{\frac{n^c}{e} + \frac{3}{2} \ln n + \ln \lambda}{rn \ln n}}) \\
 &= p_{cr} + e^{-\frac{\alpha}{r}} [1 - (1 + O(\frac{1}{n^{1-c} \ln n}))] \\
 (4.33) \quad &= p_{cr} - \Theta(\frac{1}{n^{1-c} \ln n})
 \end{aligned}$$

Thus the results are as follows:

$$Pr(Sat) > 1 - \delta, \quad \text{when } p < p_{cr} - \Theta(\frac{1}{n^{1-\varepsilon} \ln n});$$

$$Pr(Sat) < \delta, \quad \text{when } p > p_{cr} + \Theta(\frac{1}{n \ln n}),$$

where $\varepsilon = \frac{c+|c|}{2}$, $c < 1$ and $0 \leq \varepsilon < 1$.

Therefore the scaling window of model RB with respect to parameter p is

$$W(n, \delta) = (p_{cr} - \Theta(\frac{1}{n^{1-\varepsilon} \ln n}), p_{cr} + \Theta(\frac{1}{n \ln n}))$$

Remark 4.2 Similar to Remark 4.1, by Lemma 4.2, we obtain that the convergence rate of $p_-(n, \delta)$ approaching p_{cr} increases with both r and α .

Note that especially, when $n \rightarrow \infty$, we have

$$\begin{aligned} Pr(Sat) &\rightarrow 0, \text{ when } r > r_{cr} \text{ or } p > p_{cr}, \\ Pr(Sat) &\rightarrow 1, \text{ when } r < r_{cr} \text{ or } p < p_{cr}. \end{aligned}$$

This is the result of Xu and Li[16].

5. CONCLUSIONS

In this paper, we obtain the scaling window of model RB for which the phase transition point is known exactly. As mentioned before, the scaling window of random 2-SAT has also been determined. However, this model is easy to solve because 2-SAT is in P class. Recently, both theoretical[17] and experimental results[15] suggest that model RB is abundant with hard instances which are useful both for evaluating the performance of algorithms and for understanding the nature of hard problems. As far as we know, this paper is the first study on the scaling window of such a model with hard instances. We hope that it can help us to gain a better understanding of the phase transition phenomenon in NP-complete problems.

REFERENCES

1. D. Achlioptas and G. Sorkin, Optimal myopic algorithms for random 3-SAT, *Proceedings of the 41st Annual IEEE Symposium on Foundations of Computing* (2000) 590-600.
2. B. Bollobás, Christian Borgs, Jennifer Chayes, J.H. Kim, and D.B. Wilson, The scaling window of the 2-SAT transition, *Random Structures and Algorithms*, 201-256(2001).
3. P. Cheeseman, B. Kanefsky and W. Taylor, Where the really hard problems are, *Proc. 12th Int. Joint Conf. on Artificial Intelligence*, 331-337(1991).
4. V. Chvátal and B. Reed, Mick gets some (the odds are on his side), *Proc. 33rd Symposium on the foundations of Computer Science*, 620-627(1992).
5. S.A. Cook. The complexity of theorem-proving procedures, *Proc.3rd ACM Symposium on Theory of Computing*, 151-158(1971).
6. Y. Fu and P.W. Anderson, Application of statistical mechanics to NP-complete problems in combinatorial optimisation, *J. Phys. A*, 19:1605-1620(1986).
7. I.P. Gent, E. MacIntyre, P. Prosser and T. Walsh, The constrainedness of search, *In Proceedings of the 13th National Conference on AI*, pages 315-320. American Association for Artificial Intelligence, (1997).
8. A. Goerdt, A threshold for unsatisfiability, *Journal of Computer and System Sciences*, 33:469-486(1996).
9. T. Hogg, B. A. Huberman, and C. Williams, Eds., Frontiers in problem solving: phase transitions and complexity, *Artificial Intelligence* 81, 1996.
10. L. M. Kirousis, E. Kranakis, D. Krizanc and Y. C. Stamatiou, Approximating the unsatisfiability threshold of random formulae, *Random Structures and algorithms* 12(1998) 253-269.
11. S. Kirkpatrick and B. Selman, Critical behavior in the satisfiability of random boolean expressions, *Science*, 264:1297-1302(1994).
12. R. Monasson, R. Zecchina, S. Kirkpatrick, B. Selman, and L. Troyansky, Determining computational complexity from characteristic phase transitions, *Nature*, 400:133-137(1999).

13. B. Selman, H. Levesque and D. Mitchell, Hard and easy distributions of SAT problems, *Proc. 10th Nat. Conf. on Artificial Intelligence*, 459-465(1992).
14. Y. Verhoeven, Random 2-SAT and unsatisfiability, *Information Processing Letters*, 72:119-123(1999).
15. K. Xu, F. Boussemart, F. Hemery and C. Lecoutre, Random Constraint Satisfaction: Easy Generation of Hard (Satisfiable) Instances, *Artificial Intelligence*, 171(2007):514-534, Earlier version appeared in Proc. of 19th IJCAI, pp.337-342, Scotland, 2005.
16. K. Xu and W. Li, Exact phase transition in random constraint satisfaction problems, *Journal of Artificial Intelligence Research*, 12:93-103(2000).
17. K. Xu and W. Li, Many hard examples in exact phase transitions, *Theoretical Computer Science*, 355:291-302(2006).

¹SCHOOL OF SCIENCE, BEIJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS, BEIJING 100083, CHINA, KEY LABORATORY OF MATHEMATICS, INFORMATICS AND BEHAVIORAL SEMANTICS, MINISTRY OF EDUCATION

E-mail address: `xiaoyanzi@ss.buaa.edu.cn`

²SCHOOL OF COMPUTERS, BEIJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS, BEIJING 100083, CHINA, NATIONAL LABORATORY OF SOFTWARE DEVELOPMENT ENVIRONMENT

E-mail address: `kexu@nlsde.buaa.edu.cn`

³SCHOOL OF SCIENCE, BEIJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS, BEIJING 100083, CHINA, KEY LABORATORY OF MATHEMATICS, INFORMATICS AND BEHAVIORAL SEMANTICS, MINISTRY OF EDUCATION

E-mail address: `zzheng@pku.edu.cn`